熊大教育実践研究　第27号，11－18，2010

Using Formative Assessment to Improve the Teaching of English to Young Learners

Stan PEDERSON

Abstract

Formative assessment can be used to shed light on any aspect of teaching or learning as a basis for improving instruction. This paper provides examples of formative assessment in English classes for young learners to show a diversity of approaches. It follows assessment through the instructional cycle from planning, to teaching, to gathering data, interpreting results and finally to re-planning. It concludes with discussion and recommendations for implementing assessment in the newly mandated Foreign Language Activities for grades 5 and 6 in Japanese elementary schools.

1. Introduction

Children are under construction. Each day they grow and change, increasing their physical size, dexterity and strength. But inside their heads, a deeper transformation is taking place, not just in an expanding store of knowledge and experience but in the very way they look at and think about the world. In a sense, the teacher is confronted with new children and a different class every day.

In order for teachers to adapt their methods to these ever-changing pupils, they need a steady flow of fresh information about who these children are and how they are responding to instruction. Gathering, interpreting and applying this information is called "formative assessment." It is called "formative" because it serves as a guide in forming, or shaping instruction. Teachers have always done this willy-nilly, making mental notes about the proclivities and progress of their students. However these ephemeral impressions are not always accurate. Making this process conscious, consistent and systematic is a major purpose of formative assessment.

This paper will detail examples of assessment-in-action taken from my own lesson notes from elementary school classes in Tokyo. I will begin by reviewing relevant literature, follow with examples showing how formative assessment leads to improved instruction, then conclude with a discussion of how it might be used to enhance teaching in the newly mandated Foreign Language Activities in Japanese elementary schools.

2. Literature review

Formative assessment fits within the broader purposes and uses of assessment for young learners. As outlined by Cameron (2001):

> The process and outcomes of assessment can help motivate learners
> An assessment activity can provide a useful model of language use.
> [O]utcomes of assessment can help teachers plan more effective lessons.
> [O]utcomes of assessment can inform the evaluation and improvement of courses and programs. (p. 110)

Formative assessment, being more closely related to immediate teaching concerns such as making on-the-spot changes during class or changing course for the next lesson or unit, requires results to be available in a timely fashion. Not only teachers but students benefit most from immediate feedback about their performance, as "the sooner both learners and teachers have access to feedback the better so that teacher intervention and learner action can be planned." (Rea-Dickens & Germaine,1992, p. 44)

Formative assessment is typically linked to classroom activity if not embedded in the activity itself because : "It is fairer to assess children on the basis of what they have been taught and how, using assessment activities that are familiar to children from their classroom experience." (Cameron, 2001, p. 110) Not only does this test what has been taught (a validity issue), it also dampens the noise that might be created by disorienting children with a novel testing method (a matter of reliability). Children can easily be spooked by a testing method outside of their limited life experience. An assessment within the bounds of everyday classroom activity allows students to feel secure and to show the full extent of their ability.

Clay (1993, cited in Ross 2004) describes such a classroom activity based assessment, "running records", a reading assessment administered by the teacher to individual students :

> As the child reads, the teacher codes each word, reporting the percentage of words correctly read ; the self-correction ratio (the ratio of errors + self-corrections) divided by the number of self-corrections); and the categories of errors made (meaning, visual or structure). After the reading, the student retells the story and answers questions about the story's meaning. A running record is successfully completed when the student has responded correctly to the questions about meaning and has read at 90-94% accuracy. The level of the passage read and the types of errors made by the student guide subsequent instruction. (p. 187)

Such an assessment uses extant study materials and tests students performing the target activity. Because it is tailored to the level of each student, subsequent instruction can be targeted specifically to their needs.

But formative assessments may also be designed for use program-wide with common standards and with global controls on content difficulty. An example is Curriculum-based Measurement (CBM) a set of techniques developed largely in the context of special education. CBM uses pre-determined testing standards and procedures but "with stimulus materials drawn directly from the instructional materials used by teachers in their classrooms." (Deno 2003, p. 184) Measures are based on direct observation of students performing and multiple samples on equivalent tasks are taken over time to gauge student progress. Validity is assured through use of actual classroom materials and, reliability is bolstered through the multiple sampling strategy.

Reliability is exemplified in instruments like the balance used to measure mass, which shows consistent results regardless of location. The same results will be yielded in any gravitational environment. It is, in other words, context-neutral. But please note that this neutrality is not achieved by being independent from the environment but is achieved, on the contrary, by being dependent on the environment in that each side of the balance, the assessor and the assessed, share the same gravitational influence. Similarly, in assessment for young learners, reliability requires, not the dispassionate observer independent of the context, but rather the opposite ; that the teacher, as evaluator, be grounded in the knowledge of the individual development of each child within her environment. This is consistent with the general philosophy of elementary school education where "it is believed that the characteristics of each child should be known and that the child should be taught with knowledge of the combination of all these characteristics (Jalongo 2000 : Kershner 2000 as cited in McKay, 2006, p. 15) Moreover, the child should be understood in interaction with her environment. From the constructivist point of view, the child is not a repository of knowledge but rather a co-constructor of an ever-evolving reality as "children interact with others and also with their environment and concrete materials within it" (McKay 2005 p. 17) This requires that an accurate assessment capture this process *in vivo* rather than merely recording the artifacts of past experience obtained in sterile exam conditions.

The approach most thoroughly grounded in children and their environment is perhaps "pedagogical documentation" which is based in directly observing and recording incidents of what children say and do as they interact. This record then serves as a basis for reflection and discussion. MacDonald (2006) researched the use of documentation panels--composed of photographs combined with quotes from children--as a basis for shared reflection among teachers, children and parents to create a deeply grounded view of the state of teaching and learning.

However, formative assessment is not just about students and the quality of instruction. Better teaching is created through the systematic experimentation and reflection of teachers, which means that formative assessment is also a path to professional development. Frey and Fisher (2009) describe a program using "common formative assessment", that is, formative assessment developed and implemented by a group of teachers to improve program-wide results in writing instruction. The program began with teachers exchanging samples of student work. Differences were noted and questions were asked about why certain classes had more success in some areas. From these *ad hoc* beginnings the program developed a more formal approach :

> Building on their success in looking at student writing, a group of teachers met with the reading specialist and discussed the idea of creating common prompts and formative assessments so that teachers at each grade level could examine student performance data, talk about what was working, develop plans for areas needing improvement, and write curriculum that would impact their students. (p. 675)

The program resulted, not only in "impressive growth" in the Academic Performance Index for students but also "supported personal and professional development among the teachers" (p. 679) in enhancing their ability to 1) understand content demands at each grade level, 2) design assessment overall 3) link assessment with instruction, and 4) identify students with special needs.

In summary, formative assessment is used to improve classroom practice at the scale of the individual student and the whole class. It also informs pedagogy at the level of classroom decision, lesson plan and all the way up to program design. Furthermore, as opposed to more traditional styles of assessment, it requires groundedness in the child and her environment as well as in the curriculum, materials and methods of instruction. Finally, in the process of improving classroom instruction, it also raises the quality of teachers as they design assessment, reflect on the results and implement improvements.

3. Examples of formative assessment

3.1 Assessment with contingent support

Because the primary purpose of formative assessment is pedagogical, it allows the teacher to support and guide the student as part of the assessment itself. This fits in with the child's experience of the world. When a child has difficulty, they are used to being provided with guidance and assistance (contingent support) by parents, siblings or friends. Formative assessment can be used on-the-spot as an opportunity for instruction. The first example shows how this can be done and what results it can yield.

I generally greet my classes with "How are you" and receive the answer : "I'm fine." One day, however, I greeted several students individually and was surprised to hear them leaving out the /m/ in "I'm" and the /n/ in "fine." The problem demanded attention so I gave detailed instruction to make students aware of what exactly had been left out. I pressed my finger to my lips to show the bilabial formation of /m/ and I used the Japanese *hiragana* letter approximating /n/ to help them become aware of the terminal phoneme in "fine." In addition I used an illustration on the board to show the tongue placement for the English /n/ which differs from the Japanese equivalent. We practiced as a class several times. At the end of class I tested each child as they left the classroom Children who had a problem were instructed on-the-spot and given the chance to try again.

Table 1: Results of Whole-class Instruction and One-to-one Follow-up for "I'm fine" at School Y

	Whole-class instruction		One-to-one follow up	
Class	/m/ OK	/m/ missing	/m/ OK	/m/ missing
3-2	27	6	5	1
3-1	24	10	10	0
	/n/ OK	/n/ missing	/n/ OK	/n/ missing
3-2	24	9	8	1
3-1	22	12	9	3
Totals	97	37	32	5

The results (Table 1) show that in 37 of 97 or 38% of cases overall, there remained a pronunciation problem, even after extended demonstration and explanation. However, of these 37 cases, 32 were able to correct after individual on-the-spot instruction.

I interpreted this as showing : first, that whole class instruction had failed to get a positive result in about two of five students and second, that most of these students could succeed given individual attention. Not only did this assessment clarify and quantify a problem, it produced a simple and effective solution. Thereafter, I was sure to check pronunciation individually for the chance to give this one-to-one instruction. The assessment may also have been more accurate than one in which support is withheld. From a constructivist point of view, an assessment offering support reveals not only what a student can do but, more importantly, how they respond to instruction.

Note that, in this instance, I became little wiser as to the source of student problems in the whole-class situation. I do not know if my explanation was too elaborate (a known contributor to bad performance), if students in the back of the room performed more poorly than students in the front (suggesting difficulty seeing or hearing) or if the students having difficulty with this had difficulty with language in general. In other words, I was not able to isolate the variables and apportion their relative contributions to the results as one would hope to do in experimental research. I was, however, able to identify a problem and discover a way to enhance the teaching of pronunciation in future, a major goal of formative assessment.

3.2 Assessing "motivation"

When students will not participate in activities, follow directions, or do their homework it is often characterized as a lack of motivation. Unfortunately, as soon as we class the problem in such a way, we are then left with the vague problem of how to improve student motivation. It is far better to stick with the original concrete problem--students not handing in homework, for example--and see if we can do something about that. Until we specify the problem in such a clear measurable way and collect basic data, we can never know whether anything we try has worked. The example that follows concerns the specific problem of how to increase the number of children singing a song.

This assessment concerns two song activities : "London Bridge" (traditional) and "Little Mouse" (Suda & Pederson 2004). Each activity consists of an unchanging song accompanied by a game. The goal is for children to become accustomed to the song while playing the game and to join in the singing when they feel ready. While the approach seemed to have been satisfactory in the past, with most children joining in spontaneously after a time, I noticed at one school that even after a couple of lessons, few children were joining in the singing. It was, therefore, decided to find out the extent of the problem and try to increase participation.

The class was managed by an English speaking specialist (myself) and the regular classroom teacher. To establish a baseline, the classroom teacher asked the children to sing as they played the game. The number of children singing was recorded surreptitiously. In the next round children were told that their classroom teacher would be watching to see if they were singing. She held a clipboard and recorded the number of children singing as they passed in the circle. The result was announced and the game was played again. The

teacher reinforced the singing behavior with praise when she made the announcement. She also set the goal for more children to sing in the next round. The procedure was repeated once more.

There were four sets of data. The first was the baseline, recorded secretly after a general encouragement to sing. The second was the number of children singing after being made aware of being observed and recorded for singing behavior. The third and fourth sets included the element of class feedback about how many kids were singing but also included general praise and setting the task of more pupils singing.

Table 2: London Bridge and Little Mouse:Singing in Three conditions:School X Grade Ones, 2007

Treatment	"London Bridge"				"Little Mouse"					
	1-1		1-2		1-1		1-2		Totals	
	Sing	No	Sing	No	Sing	No	Sing	No	Sing	No
Baseline	15	9	12	13	20	3	13	9	60 (64%)	34 (36%)
Observed	21	3	19	4	23	0	18	4	81 (88%)	11 (12%)
Feedback 1	23	1	23	2	23	0	20	2	89 (95%)	5 (5%)
Feedback 2	*	*	25	0	23	0	20	2	68 (97%)	2 (3%)

*No fourth round in this class

It seems that being blatantly assessed (being alerted to the assessment and brandishing the clipboard) had a powerful influence on the participation of these grade ones. They moved from 64% singing to 88% singing. Feedback and praise combined with setting a goal had a further impact pushing the participation rate up to 95% (feedback 1) and 97% (feedback 2).

As in the pronunciation example, disambiguating the influence of different variables is impossible. This method fails to isolate the relative influence of blatant observation and feedback, praise, encouragement and the act of specifying what was being assessed. It should also be noted that the classroom teacher most probably praised singing students with her eyes or with a smile as they passed by. While it fails to meet the requirements of an experiment, this set of procedures used together did, nonetheless, produce an improved outcome and can be added as a whole to future teaching plans. A simple count of observable behavior during an activity gave us an accurate and objective picture of the real situation in addition to providing a baseline against which to measure the effectiveness of a cluster of pedagogical interventions. The assessment involved zero cost in off-task activity and resulted in changes to the teaching of similar activities for grade ones in the future.

3.3 Assessing adherence to instructions

The third example assesses whether students follow instructions in an activity called "How many monkeys in the tree?" (Suda & Pederson 2003). In this activity students compete pair against pair, alternately asking a question then making a guess about the number of monkeys on the other team's randomly drawn card. Three rules are imposed for pedagogical reasons. First, students are asked to chant the question together, meaning the stronger student provides a model for the weaker student while the weaker student experiences both support and pressure to perform more quickly and accurately. Second, they are to consult each other on the number to be guessed and say it together after reaching agreement. This produces real communication and negotiation around a simple decision. And third, they are to alternate turns with the opposing team—asking, then guessing to maintain active involvement of all members for the duration of the game. Any departure from these rules weakens the game as a learning activity.

After noticing a number of difficulties I decided to find out the extent of the problem. After a few lessons it was discovered that showing the CD-ROM demonstration and pausing for a pointer to alternate turns had eliminated this last problem completely so it was dropped from the assessment protocol. On the other hand, it was also noticed that students often failed to pronounce the final /z/ sound in "monkeys" so this was added to the protocol used by teachers as they assessed each group.

Table 3 shows the results for one grade-two class doing the game for the first time. The target expressions were reviewed. The game was then shown on CD-ROM while pointing out teams alternating turns, pairs consulting with their partner and chanting questions together.

Table 3: Assessment of "How Many Monkeys in the Tree" Activity:Class 2-2 School X , 2007

	Chant together	Consult partner	Monkey/z/
Pair 1	yes	yes	yes
Pair 2	yes	yes	no
Pair 3	yes	yes	?**
Pair 4	yes	yes	yes
Pair 5	no	yes	?
Pair 6	no	alternate*	?
Pair 7	yes	no	yes
Pair 8	yes	yes	yes
Pair 9	no	yes	yes
Pair 10	yes	alternate	yes

*Pair 6 and 10 alternated which partner guessed the answer.
**Not audible to assessor

Fleshing out my post lesson notes I can say that I thought not chanting together to be a concern. Perhaps the children needed more work in producing the form before trying the game again next time. Three pairs did not consult their partner before guessing. Two pairs alternated who decides which number to guess. This got the decision made but it short-circuited the necessity to negotiate, an important goal of the activity. Consulting with their partner to produce an agreed guess will need to be restated next time. Lastly, the question marks produced in the column regarding the terminal sound in "monkeys" point to it being ineffective to try to assess this during a pair chant. It might be better to check a sample of students, perhaps one or two rows, before the activity begins and point out successes or offer corrections in front of the class.

After some trial and error with various classes, the final sound on "monkeys" was dealt with prior to the game by sampling two or three rows of students. Problems with following the rules more or less disappeared when we added a student demonstration of the game combined with pointers in Japanese. The assessment was discontinued once it had solved the problems.

3.4 Assessment as exploring

The final example involves the same activity but shows an exploratory approach to assessment. In this case, since rules were being followed and no major structural problems were evident, we found a new question to investigate : which numbers are the children guessing? Each teacher recorded the numbers guessed as they circulated among the groups. Numbers one to three were not included in the game as children had mastered them previously. The results appear in Table 4.

Table 4: Count of Numbers Guessed

	four	five	six	seven	eight	nine	ten	eleven	twelve
Game 1	4	8	5	6	6	2	4	4	1
Game 2	7	3	9	8	7	5	4	3	4
Total	11	11	14	14	13	7	8	7	5

While the results of one game do not seem to reveal any trends, it is clear with the combined observations of two games that children are guessing 4-8 more frequently than they are guessing 9-12. Since guessing more lower numbers is of no advantage in a random guessing game, we interpreted this as the children tending to avoid numbers they were less able to recall or produce. In other words, this set of observations seemed to reveal something about the level of vocabulary learning. This led to reviewing only numbers 9-12 in the next lesson. Such an assessment took no extra class time and allowed us to focus only on what the children needed

to review, thus saving time in the next class. More importantly, it suggested a way to assess vocabulary learning in any other game where children may choose freely among vocabulary items.

4. Discussion

In the examples above, I have dealt with assessment as a tool for improving teaching. This is the essence of formative assessment. However, for many, assessment means assigning a performance level to individual students, and assigning grades to students is, in fact, required in most programs. But there need be no conflict between the two sorts of assessment for they differ mainly in how the data is presented and with whom it is shared. If you must assign grades or performance descriptors, the data can be used to focus on individual students.

In the pronunciation example, for instance, we can assign three levels of performance : 1) able to pick up English pronunciation demonstrated and explained in class, 2) able to pick up English pronunciation given individual instruction, and 3) needs more work on fundamentals of English pronunciation.

In the second example about joining in with singing we can also define three levels : 1) joins in singing English songs spontaneously, 2) joins in singing when encouraged by the teacher, and 3) participates in game but is not yet able to sing at the same time.

In the third example regarding chanting together and consulting with their partner we can use a few simple descriptors to evaluate behavior. For example, we might say : is able to express an opinion and listen to the opinions of others : is able to chant with classmates and ; follows instructions in classroom activities.

Under the new course of study, which makes Foreign Language Activities compulsory for grades five and six, traditional summative assessment, using formal tests, ranking students against each other and assigning grades is not permitted. Some have interpreted this to mean that there is no assessment. This is certainly not the case and it is made clear in the teacher's manual of the MEXT produced Eigo Noto, a guidebook of model lessons and activities.

Eigo Noto makes clear that evaluation is to be conducted in reference to the three general aims, or pillars, of Foreign Language Activities :

1. By way of foreign language and through experiences to deepen understanding of foreign language and culture.
2. By way of foreign language, to foster an attitude of willingness to communicate.
3. Through the medium of foreign language, to develop familiarity with the sounds of the foreign language and some of its fundamental expressions. (Author's translation) *Eigo Notto* : Shidou Shiryou, Dai 5 Gakunen Shikouban p. 8

It proceeds to give 8-9 examples of assessments within each category. 23 examples are based on observing children during activities with one being supplemented by checking the students' books. The remaining 2 instances are observing student presentations. It adds a note condoning both self and peer assessment with the example of evaluation criteria being set by students themselves. The section ends with a final caution that it is not necessary to determine whether an activity has resulted in the children having acquired a particular expression.

While the emphasis on classroom observation is not misplaced, implementing the observations is crippled by a lack of behavior descriptors. Furthermore, while goals are well aimed, they are often loose and contain vague adjectives. We can, therefore, gain much from incorporating concrete techniques like those above to clarify observations and make them systematic. For example, one of the goals to be assessed by observation is : While observing manners, eagerly impart greetings (p. 8) (Author's translation). "Manners" may seem vague but manners like maintaining eye contact are language/culture specific and are clearly identified in the lesson. This means the teacher can directly observe such behaviors in the classroom : so far, so good. The

problem comes with the overused adjective "*sekkyokuteki ni*," (positively, eagerly). It is not easy to understand what this means. How would you know it if you saw it?

There are many ways to make this concrete. Some teachers would equate this adjective with the intensity of behavior as expressed in loudness of voice. In a sense this may seem no clearer than the adjective it is trying to operationalize. After all, what is loud for one student may be quiet for another. However, given the experience of the classroom teacher, seeing each student every day, she may well be in a good position to make a judgment about whether they are speaking less loudly, more loudly or at the same volume given a similar situation in their native language.

Another teacher might equate "eagerly" with the frequency of the target behavior : How many times do they perform a greeting in a given period of time? For example, how many students do they exchange greetings with in a three minute activity? This would be easy to implement in a whole class if each student records their own total. The teacher could elicit and record this information after the activity, using it as listening and speaking practice. Recording such information would assist in setting benchmarks for acceptable levels of performance in future classes and in assessing teaching interventions.

A possibility for exploration is to observe the greetings before and after the target greetings. When students form a new pair, they will probably engage in some sort of greeting ritual before moving into the target greeting activity. Do they smile? Do they touch? Do they speak in the native language or the target language? After careful observation and consideration, the teacher could choose any of these as an indication of greeting "positively/eagerly" and create a construct with rankings of descriptors along a continuum.

Or, perhaps "positively/eagerly" refers to how the student feels about engaging in greetings in English. In that case, a teacher might utilize survey style questions using scales or comments in which students would indicate whether they want to do the activity again, or how willing they would be to use such greetings in a real situation. Whichever way the teacher chooses to operationalize the adjective "positively/eagerly", she will now have something specific to observe and record and she will be able to judge when there is more or less of it.

Converting the vague goals and adjectives of *Eigo Noto* into observable behaviors could give teachers a valuable starting point for improving their classes. This would not only help them gauge the progress of their students and the effectiveness of their teaching methods but it might also give them the confidence to explore alternatives since they will always be able to determine their own direction and to understand whether they are making progress. Incorporating formative assessment in this way would allow teachers to apply their store of existing knowledge about education and about their students and to make English lessons their own.

References

Calfee, R., Hiebert, E. (1991). Classroom assessment of reading. In R. Barr, M.Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of research on reading.* (2nd ed., pp. 281-309). Longman : New York.

Cameron, L. (2001). Teaching languages to young learners. Cambridge University Press : Cambridge.

Deno, S. (2003). Developments in curriculum-based measurement. *The Journal of Special Education,* 37(3) 184-92.

Frey, N., Fisher, D. (2009). Using common formative assessments as a source of professional development in an urban American elementary school. *Teaching and Teacher Education,* 25 674-80.

MacDonald, M. (2007). Toward formative assessment : The use of pedagogical documentation in early elementary classrooms. *Early Childhood Research Quarterly,* 22 232-42.

McKay, P. (2006). *Assessing young language learners.* Cambridge University Press : Cambridge

Rea-Dickens, P., Germaine, K. (1992). *Evaluation.* Oxford University Press : Oxford

Ross, J. (2004). Effects of running records assessment on early literacy achievement. *The Journal of Educational Research,*97(4)186-194.

Suda, M., Pederson, S. (2004). Rondon Bashi [London Bridge]. *English stadium : Popperstep Kids* (Level 1): Benesse Corporation : Tokyo.

Suda, M., Pederson, S. (2004). Konezumi Gemu [Little Mouse]. *English stadium : Popperstep Kids* (Level 1): Benesse Corporation : Tokyo.